

Fill in the blanks:

6. A data mining query is defined in terms of _____.
7. A graphical method for summarizing the distribution of given attributes is _____.
8. A transaction that does not contain any frequent k-itemsets cannot contain any frequent _____ itemsets.
9. _____ removes subtrees from a “fully grown” tree.
10. _____ is a generalization of both Euclidean distance and Manhattan distance.

State True or False:

11. Tight coupling means that a DM system is smoothly integrated into the DB/DW system.
12. A unique rule says that there is no missing value between the lowest and highest values for the attributes.
13. Each level of abstraction has its own minimum support threshold.
14. Bayesian classifiers are also useful in that they provide a theoretical justification for other classifiers.
15. A cardinal variable is a generalization of the binary variable in that it can take on more than two states.

Match the following:

16. Heterogeneous database - a) The join step
17. Min-max normalization - b) Legacy database
18. Apriori algorithm - c) Linear transformation
19. Bayesian network - d) Cosine measure
20. Vector objects - e) Belief network

SECTION – B

Answer all the questions:

5 x 7 = 35

21. a. Give a detailed description about Relational Database with an example.

OR

- b. With a diagram and example explain the Classification and prediction in Data Mining.
22. a. Write short notes about the Attribute subset Selection.

OR

- b. Explain the concept of histogram with a suitable example.
23. a. Explain the constraint based mining along with its types.

OR

- b. How the Association rules are used in Quantitative Mining?
24. a. Explain how IF-THEN rules are used in Rule Based Classification.

OR

- b. With a neat diagram explain the architecture of a Multilayer Feed-Forward Neural Network.
25. a. What are the typical requirements of clustering in data mining?

OR

- b. Explain the BIRCH Hierarchical Method.

SECTION – C

Answer any THREE questions:

3 x 15 = 45

26. Describe the major issues in Data Mining.

27. Give a short description about the method used to discretize hierarchy generations for numerical data.
28. Explain how the Apriori Algorithm is used to find frequent itemsets using candidate generation.
29. Describe in detail about the back propagation learning algorithm along with its pseudo code.
30. Describe the categorization of Major Clustering Methods.
